
Instructions

Pittawat 6304691134

- (1) Please read the instruction carefully. Also take this habit with you into the exam room.
- (2) Please read each question carefully and answer the questions straightforwardly. Always provide economic reasons at least a paragraph for your analysis, or a graph when necessary, even when the question does not indicate so.
- (3) Handing and submitting assignments are only available via BE Moodle.

Answering the questions and preparing answer sheets

- (1) Answers are to be handwritten, in either digital or analog form, in a blank canvas or any clean paper. Make sure that your handwriting is clearly visible and readable.
- (2) There is no need to rewrite the question. Just indicate the question number clearly for each of the answer, such as 1.a).
- (3) Default decimal point is 4.
- (4) Choose precise wordings, especially when you want to interpret the meaning of a test, confidence interval, or coefficients.
- (5) When done, for the digital case, collage all the pages into a single PDF file. For those who write on sheets of paper, take photo of all pages then convert all of them into a single PDF file as well.
- (6) Name your PDF file as StudentID_YourNickname, such as 640123456_Bo.

Submitting your answers

- (1) Make sure your file does not exceed 10MB. This is the maximum file size for BE Moodle upload.
- (2) Login to BE Moodle, head into the course, then the assignment topic.
- (3) Choose your file to submit. Done. There will be timestamp for your upload date and time, so please make sure to not submit later than that.

For all questions, answer up to 4 decimal places

Question 1. (15 points) Given this information

$$\begin{aligned}
 n &= 18 & \sum_{i=1}^n X_i &= 388.00 & \sum_{i=1}^n Y_i &= 50.90 \\
 \sum_{i=1}^n (X_i)^2 &= 9,620.00 & \sum_{i=1}^n X_i Y_i &= 1,254.90 \\
 \sum_{i=1}^n (X_i - \bar{X})^2 &= 211.00 & \sum_{i=1}^n (Y_i - \bar{Y})^2 &= 2.5844 \\
 \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) &= 20.58 & \sum_{i=1}^n \hat{u}_i^2 &= 0.5781
 \end{aligned}$$

Use the above sample information to answer all the following questions. Show explicitly all formulas and calculations.

- From regression model: $Y_i = \beta_1 + \beta_2 X_i + u_i$, $u_i \sim NIID(0, \sigma^2)$, **find the estimators** of β_1 and β_2 with OLS method. Interpret the intercept and slope coefficients.
- Compute the value of R^2 and explain its meaning.
- If $X_i = 30$, estimate the value of $\hat{Y}_i$ and explain its meaning.
- Calculate the estimators of $\text{var}(u_i)$, $\text{var}(\hat{\beta}_1)$ and $\text{var}(\hat{\beta}_2)$.
- What are the 90-percent confident intervals for β_2 ? Interpret the meaning.
- Test the hypothesis whether the slope coefficients are different from zero at 0.05 level of significance.

a.)

$$Y_i = \beta_1 + \beta_2 x_i + u_i$$

$$u_i = Y_i - \beta_1 - \beta_2 x_i$$

$$u_i^2 = (Y_i - \beta_1 - \beta_2 x_i)^2$$

$$\min_{\beta_1, \beta_2} \sum u_i^2 = \sum (Y_i - \beta_1 - \beta_2 x_i)^2$$

$$\bullet \hat{\beta}_1 : \frac{\partial \sum u_i^2}{\partial \beta_1} = 0$$

$$2 \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0$$

$$\text{FOC.} \quad 2 \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) (-1) = 0$$

$$\sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0$$

$$\frac{\sum y_i}{n} - \frac{\sum \hat{\beta}_1}{n} - \frac{\sum \hat{\beta}_2 x_i}{n} = 0$$

$$\bar{y} - \hat{\beta}_1 - \hat{\beta}_2 \bar{x} = 0$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} \quad - (1)$$

$$\bullet \hat{\beta}_2 : \frac{\partial \sum u_i^2}{\partial \beta_2} = 0$$

$$2 \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2 = 0$$

$$\text{FOC} \quad 2 \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) (-x_i) = 0$$

$$\sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)(x_i) = 0 \quad - (2)$$

plug (1) into (2)

$$\sum (y_i - (\bar{y} - \hat{\beta}_2 \bar{x}) - \hat{\beta}_2 x_i)(x_i) = 0$$

$$\sum [(y_i - \bar{y}) x_i - \hat{\beta}_2 (\bar{x} - x_i) x_i] = 0$$

$$\sum (y_i - \bar{y}) x_i = \sum \hat{\beta}_2 (\bar{x} - x_i) x_i$$

$$\frac{\sum (y_i - \bar{y}) x_i}{\sum (\bar{x} - x_i) x_i} = \hat{\beta}_2$$

$$\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (\bar{x} - x_i)^2} = \hat{\beta}_2$$

$$\therefore \hat{\beta}_2 = \frac{20.58}{24} = 0.8575 \quad - \text{slope}$$

$$\therefore \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} = 2.8278 - 0.8575(2.95556) = 0.7261 \quad - \text{intercept}$$

b.)

$$R^2 = 1 - \frac{\sum u_i^2}{\sum (y_i - \bar{y})^2}$$

$$R^2 = 1 - \frac{0.5781}{0.5844}$$

$$R^2 = 0.7767$$

R^2 has high confidence because it has small range and the data closes to SRF line

c.

$$\begin{aligned}\hat{y}_i &= \hat{\beta}_1 + \hat{\beta}_2 x_i \\ &= 0.7261 + 0.0975(30) \\ &= 3.651\end{aligned}$$

When $x_i = 30$, expected value is 3.6503

d.)

$$\hat{\sigma}^2 = \frac{\sum U_i^2}{n-k} = \frac{0.5761}{18-2} = 0.0361$$

$$b_{\hat{\beta}_1}^2 = \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2} \cdot \hat{\sigma}^2 = \frac{9620}{18(711)} \cdot 0.0361 = 0.0910$$

$$b_{\hat{\beta}_2}^2 = \frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2} = \frac{0.0361}{211} = 1.71 \times 10^{-4}$$

e.)

$$P(\hat{\beta}_1 - t_{16, 0.05} (se_{\hat{\beta}_1}) \leq \beta_1 \leq \hat{\beta}_1 + t_{16, 0.05} (se_{\hat{\beta}_1})) = 1 - \alpha$$

$$\alpha = 0.1, \text{ d.f.} = 16, se_{\hat{\beta}_1} = \sqrt{1.71 \times 10^{-4}} = 0.0131, t_{16, 0.05} = 1.746, \hat{\beta}_1 = 0.0975$$

$$P(0.0975 - 1.746(0.0131) \leq \beta_1 \leq 0.0975 + 1.746(0.0131)) = 1 - \alpha$$

$$(0.0796 \leq \beta_1 \leq 0.1204)$$

$\therefore$ at 90 out of 100 times, β_1 is between 0.0796 and 0.1204

f.)

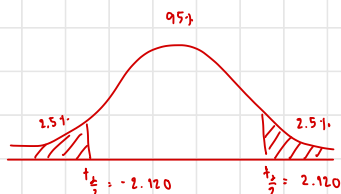
$$① H_0: \beta_2 = 0$$

$$H_1: \beta_2 \neq 0$$

$$② \alpha = 0.05$$

$$③ t = \frac{0.0975 - 0}{0.0131} = 7.4429$$

$$④ t_{\frac{\alpha}{2}} = t_{0.025} \text{ df} = 16 = \pm 2.120$$



⑤ $t_{cal} > t_{16, 0.025}$, It will be in rejection region and reject null hypothesis and conclude that as 1. CI, β_2 is not equal to zero

Question 2. Using the 2015 Health and Welfare Survey from the National Statistical Office, a simple linear regression is modeled as follows,

$$outp_i = \beta_1 + \beta_2 age_i + u_i$$

where $outp_i$ is how many times person i has visited hospital in 2015, from 0 to 7 times
 age_i is how old is person i , from 0 to 97 years.

We assume that both $outp_i$ and age_i are continuous, the estimation results in the following table. Answer the following questions and show your work.

Source	SS	df	MS	Number of obs	=	27,886
Model	77.5444409	1	77.5444409	F(1, 27884)	=	186.96
Residual	11565.0627	27,884	.414756231	Prob > F	=	0.0000
				R-squared	=	0.0067
				Adj R-squared	=	0.0066
Total	11642.6072	27,885	.417522223	Root MSE	=	.64402

outp	Coefficient	Std. err.	t	P> t	[95% conf. interval]
age	.0031338	.0002292	Omitted		.0026846 .003583
_cons	.4279898	.0140339			.4004828 .4554969

- Test if both parameters are significantly different from zero or not. Use $\alpha = 0.05$.
- Interpret the meaning of $\hat{\beta}_2$. Does the sign of $\hat{\beta}_2$ make economic sense? Explain.
- If $outp_i$ is turned into natural logarithmic scale ($\ln$), how would you reinterpret the relationship between $\hat{\beta}_2$ and $\widehat{outp}_i$, assumed that the given coefficient given in the table above can be used to interpret this new functional form.
- If age_i variable is divided by 10, how does it affect both the coefficients, standard errors, and confidence intervals? Answer the changes of both the constant and slope (if there is).
- Find the confidence interval of mean prediction at the age of 50 years old, given that $var(\hat{Y}_0) = 0.00002$ and $\alpha = 0.01$.

a.)

① $H_0 : \rho_1 = 0$

$H_1 : \rho_1 \neq 0$

② $\alpha = 0.05$

③ $t_{cal} = \frac{0.4279898}{0.0190779} - 0 = 30.4969$

④ $t_{\frac{\alpha}{2}} = t_{0.025} \text{ df } 27.885 = \pm 1.960$



⑤ t_{cal} falls into rejection region, we can reject null hypothesis at 95% CI and ρ_1 is not equal to zero

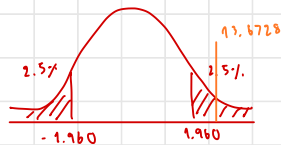
① $H_0 : \rho_2 = 0$

$H_1 : \rho_2 \neq 0$

② $\alpha = 0.05$

③ $t_{cal} = \frac{0.0031378}{0.0007192} = 13.6728$

④ $t_{\frac{\alpha}{2}} = t_{0.025} \text{ df } 27.885 = \pm 1.960$



⑤ t_{cal} falls into rejection region, we can reject null hypothesis at 95% CI and ρ_2 is not equal to zero

b.)

$\hat{\rho}_2$ make economic sense. That is because when age increase, the opportunity to visit hospital increase as well. As shown, when age increases 1-year-old, the average to visit hospital will increase by 0.0031378 time

c.)

$$\widehat{\text{out}}_{pi} = \hat{\beta}_1 + \hat{\beta}_2 \text{ age}_i$$

$$\widehat{\text{out}}_{pi} = 0.4279898 + 0.031338 \text{ age}_i$$

$$\ln \widehat{\text{out}}_{pi} = 0.4279898 + 0.031338 \text{ age}_i$$

$$\frac{\ln \widehat{\text{out}}_{pi}}{\text{age}_i} = 0.031338$$

$\therefore$ As people getting older by 1 year, $\widehat{\text{out}}_{pi}$ or Y will increase by 3.1338 %.

d.)

If age_i variable is divided by 10, the coefficients, standard errors, and CI of age will change. The constant is 0.4279898 time and slope (coefficient of age) is 0.00031338 year

e.)

$$\text{se}_{\hat{Y}_0} = \sqrt{\text{Var}(\hat{Y}_0)} = \sqrt{0.00002} = 4.4721 \times 10^{-3}$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2$$

$$\hat{Y}_i = 0.4279898 + 0.0031338 X_i$$

at $X = 50$ $\hat{Y}_{50} = 0.4279898 + 0.0031338(50)$

$$\hat{Y}_{50} = 0.5846798$$

$$\bullet \quad t_{\frac{\alpha}{2}} = t_{0.005} \quad \text{df } 27,885 = 2.576$$

$$P(\hat{Y}_0 - (t_{\frac{\alpha}{2}} \cdot \text{se}_{\hat{Y}_0}) \leq Y_0 \leq \hat{Y}_0 + (t_{\frac{\alpha}{2}} \cdot \text{se}_{\hat{Y}_0})) = 1 - \alpha$$

$$P(0.5847 - (2.576 \times 4.4721 \times 10^{-3}) \leq Y_0 \leq 0.5847 + (2.576 \times 4.4721 \times 10^{-3})) = 0.99$$

$$P(0.5772 \leq Y_0 \leq 0.5962) = 0.99$$

$\therefore$ at CI 99%, Y_0 is between 0.5772 and 0.5962

Question 3. Discuss in a short paragraph why the confidence interval for both the mean prediction and individual prediction get larger as the X_0 is further away from $\bar{X}$.

- Mean prediction

$$\hat{y}_0 = \hat{\beta}_1 + \hat{\beta}_2 x_0$$

when $\hat{y}_0$ is an estimator of $E(y|x_0)$ while x_0 represents a value of interest

- Individual prediction

$$fe = \hat{y}_0 - y_0$$

individual prediction focuses on forecasting error (fe) and we can define variance of this fe as

$$\text{var}(\hat{y}_0 - y_0) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]$$

$$\Pr \left[\hat{y}_0 - (t_{\frac{\alpha}{2}} \cdot se_{fe}) \leq y_0 \leq \hat{y}_0 + (t_{\frac{\alpha}{2}} \cdot se_{fe}) \right] = 1 - \alpha$$

when $x_0 - \bar{x}$ is larger, the variance is larger and $se(\hat{y}_0)$ also increase

$\therefore$ The big variance make data lower confident